EXHIBIT 8

US005182765A

United States Patent [19]

Ishii et al.

[11] Patent Number: 5,182,765

[45] Date of Patent: Jan. 26, 1993

[54] SPEECH RECOGNITION SYSTEM WITH AN ACCURATE RECOGNITION FUNCTION

[75] Inventors: Takaaki Ishii, Sagamihara; Toru Kuge, Hino, both of Japan

[73] Assignee: Kabushiki Kaisha Toshiba, Kawasaki, Japan

[21] Appl. No.: 825,421

[22] Filed: Jan. 24, 1992

Related U.S. Application Data

[63] Continuation of Ser. No. 601,013, Oct. 24, 1990, abandoned, which is a continuation of Ser. No. 382,346, Jul. 20, 1989, abandoned, which is a continuation of Ser. No. 933,213, Nov. 21, 1986, Pat. No. 4,873,714.

[30] **Foreign Application Priority Data**

Nov. 26, 1985 [JP] Japan .... 60-181654[U]
Nov. 26, 1985 [JP] Japan 60-265502
Dec. 20, 1985 [JP] Japan 60-287434

[51] Int. Cl.[5] G10L 9/08; H04M 1/30
[52] U.S. Cl. 379/88; 379/58; 379/63; 379/354; 381/43
[58] Field of Search 379/88, 58, 63, 354, 379/355, 216; 381/42, 43

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,348,550	9/1982	Pirz et al.	379/357
4,461,023	7/1984	Katayama	381/43
4,525,793	6/1985	Stackhouse	364/513.5
4,644,107	2/1987	Clowes et al.	379/354
4,731,811	3/1988	Dubus	379/58
4,852,171	7/1989	Jakatdar et al.	381/43
4,853,953	8/1989	Fujisaki	379/88

FOREIGN PATENT DOCUMENTS

0012463	1/1983	Japan	379/354
0059846	4/1985	Japan	379/355
974850	11/1964	United Kingdom	379/88

OTHER PUBLICATIONS

Wescon 83 Conference Record, Nov. 8th–11th 1983, San Francisco, pp. 11/4/1–11/4/3, Western Periodical Co., North Hollywood, US; D. Kelley: "Voice recognition is user-friendly" *p. 11/4/1, right-hand column, lines 19,20*.

Electrical Communication, vol. 59, No. 3, 1985, pp. 281–285, Harlow, G. B.; M. Immendorfer: "Voice Dialer" *pp. 283, 284: Features of the voice dialer; p. 283, left-hand column, lines 18-50*.

Japan Telecommunications Review, vol. 24, No. 3, Jul. 1982, pp. 267–274, Tokyo, JP; N. Ishii et al.: "Speaker-Independent speech recognition unit development for telephone line use" *p. 269, left-hand column, lines 11–22.

Primary Examiner—Thomas W. Brown
Attorney, Agent, or Firm—Oblon, Spivak, McClelland, Maier & Neustadt

[57] **ABSTRACT**

A speech registration/recognition processing section is responsive to a speech input signal to be registered or recognized from a microphone to selectively subject the speech input signal to a registration or a recognition processing, in which upon the registration processing the speech input signal is allowed to be stored as recognition data and upon the recognition processing the speech input signal is compared to the recognition data stored. A speech record/reproduction processing section is responsive to the speech input signal from the microphone to subject the speech input signal to a record/reproduction processing, in which upon the record processing the speech input signal is recorded as a record signal and upon the reproduction processing the record signal is delivered as a reproduction signal. A speaker is supplied with the reproduction signal and the recognized information is provided to the user through this speaker. A control section is provided, in accordance with a registration or a recognition mode designation signal, for setting the speech registration/recognition processing section to the registration or the recognition processing and for setting the speech record/reproduction processing section to the record or the reproduction processing corresponding to the registration or the recognition processing at which the speech registration/recognition processing section is placed.

1 Claim, 8 Drawing Sheets

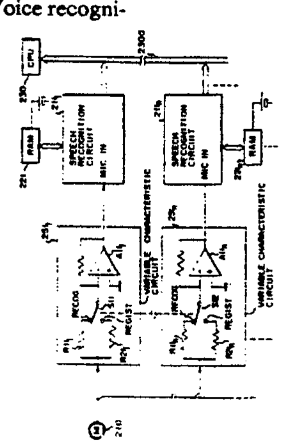
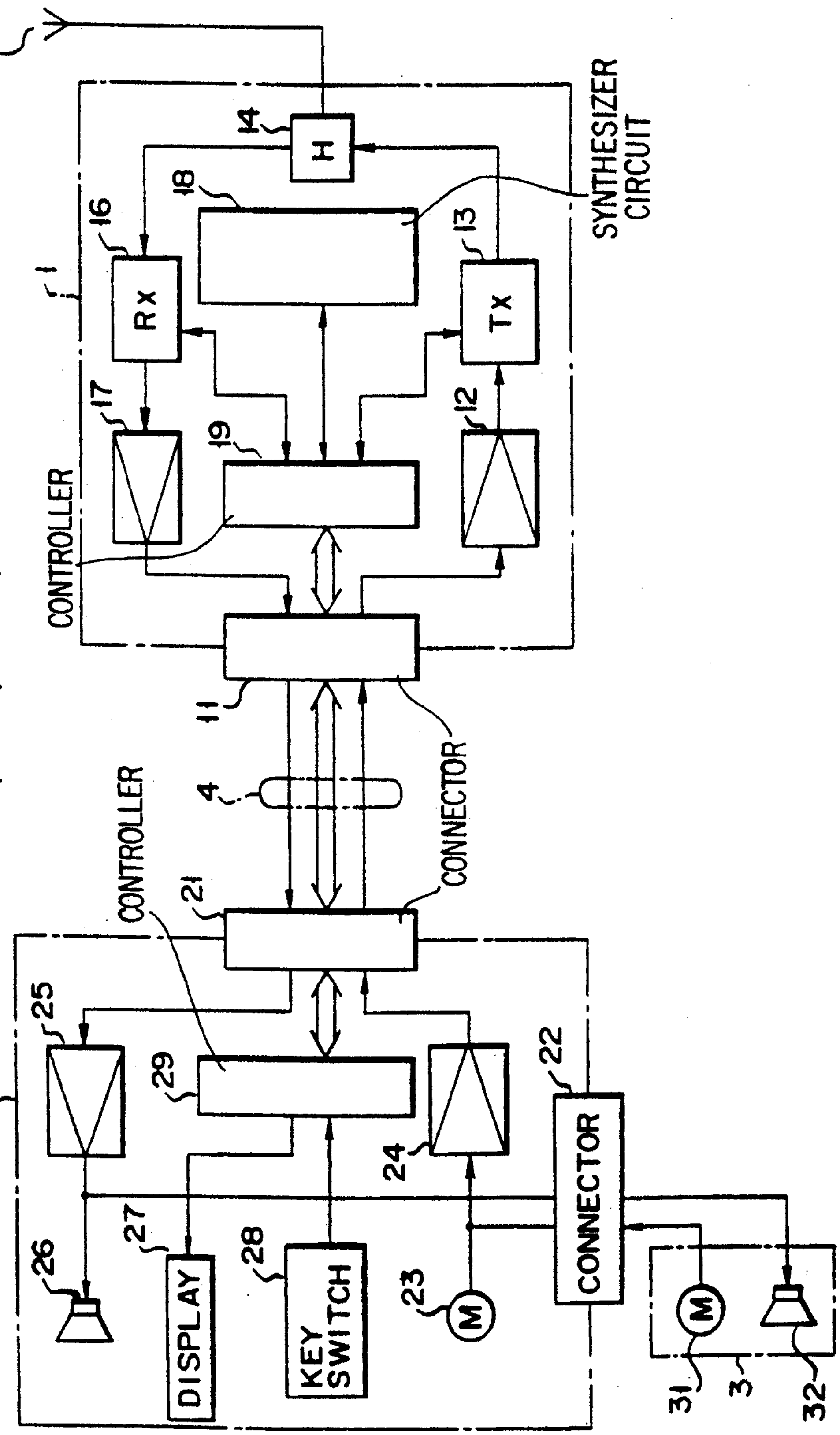


FIG. 1


15
14
H
18
SYNTHESIZER CIRCUIT
16
Rx
13
Tx
1
17
19
12
CONTROLLER
11
CONNECTOR
4
CONTROLLER
21
25
2
29
22
24
CONNECTOR
27
26
28
23
M
DISPLAY
KEY SWITCH
M
31
3
32

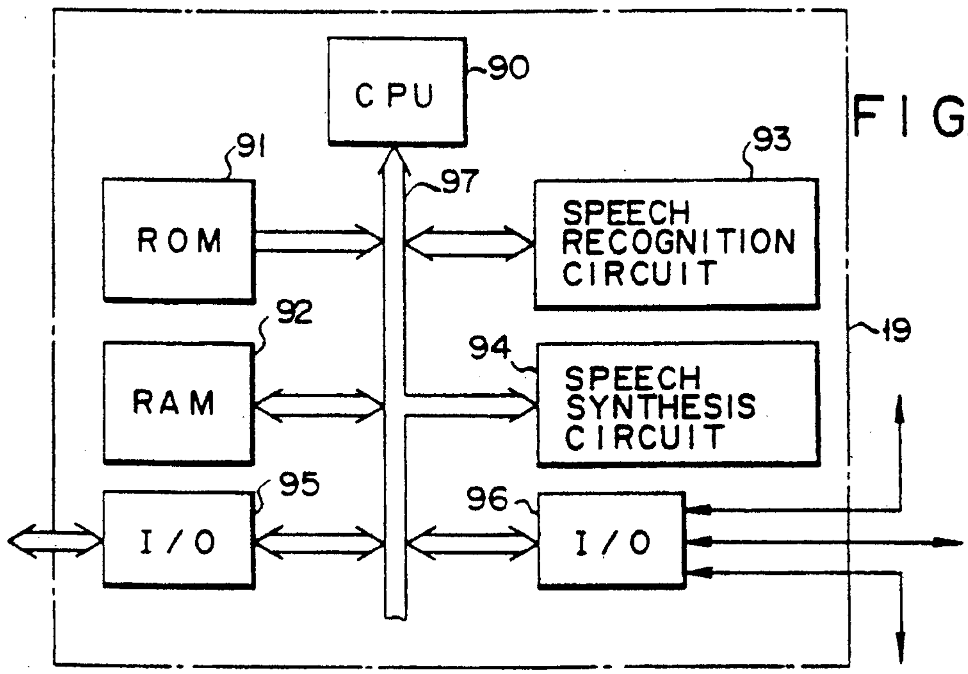
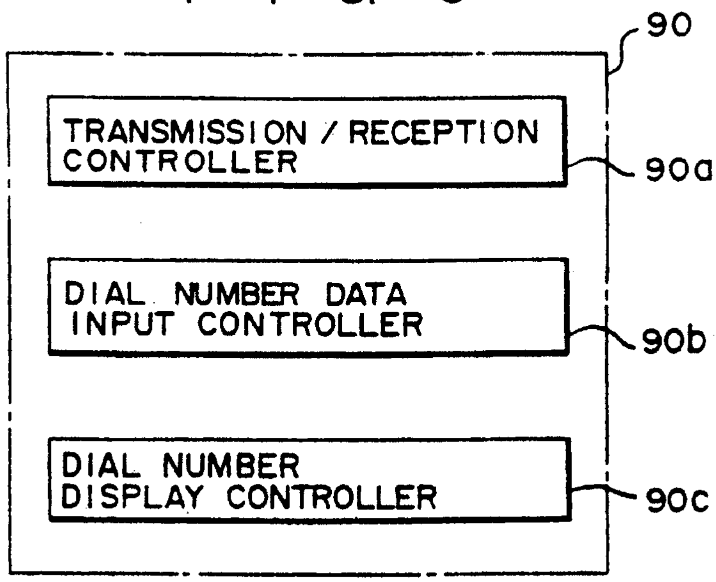

CPU
90
91
93
97
ROM
SPEECH RECOGNITION CIRCUIT
92
94
19
RAM
SPEECH SYNTHESIS CIRCUIT
95
96
I/O
I/O

FIG. 2

FIG. 3


90
TRANSMISSION / RECEPTION CONTROLLER
90a
DIAL NUMBER DATA INPUT CONTROLLER
90b
DIAL NUMBER DISPLAY CONTROLLER
90c

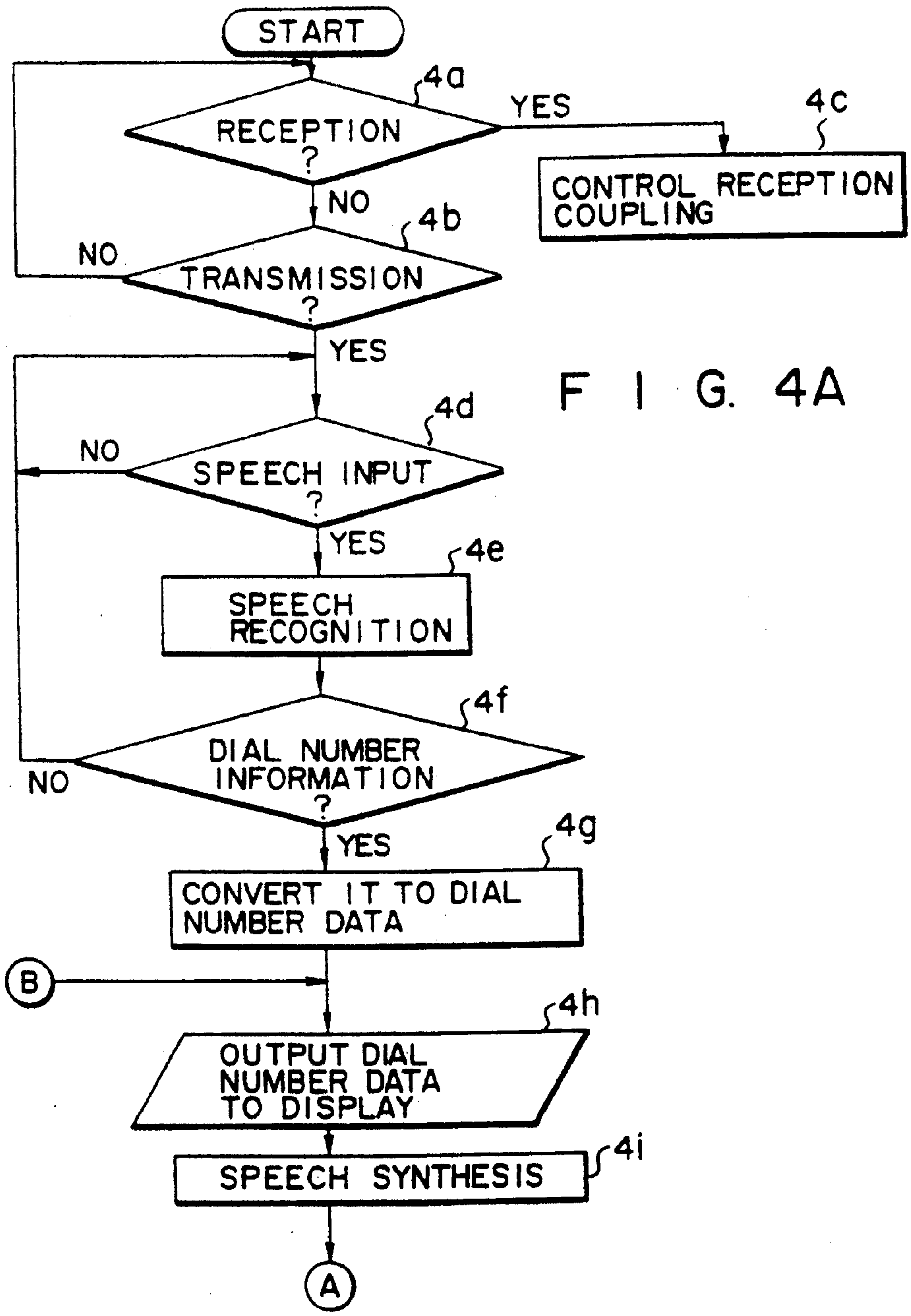
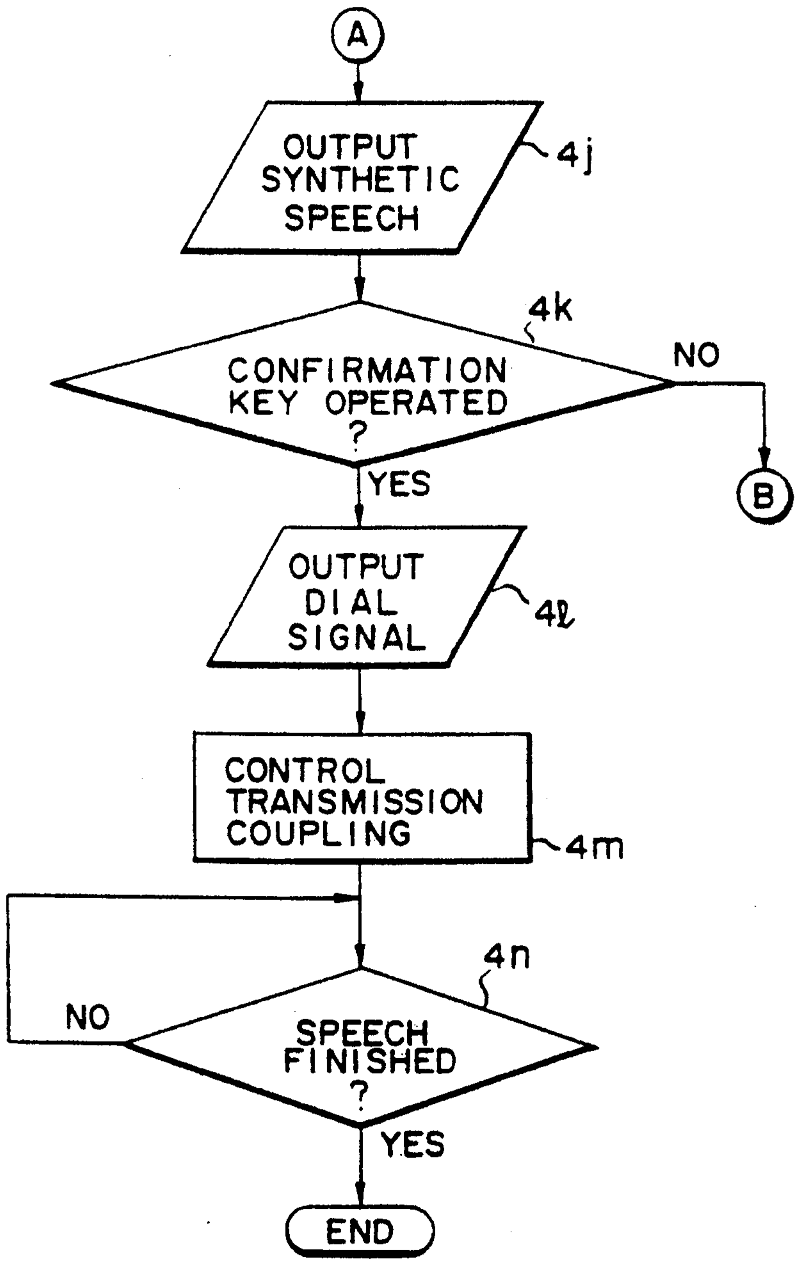
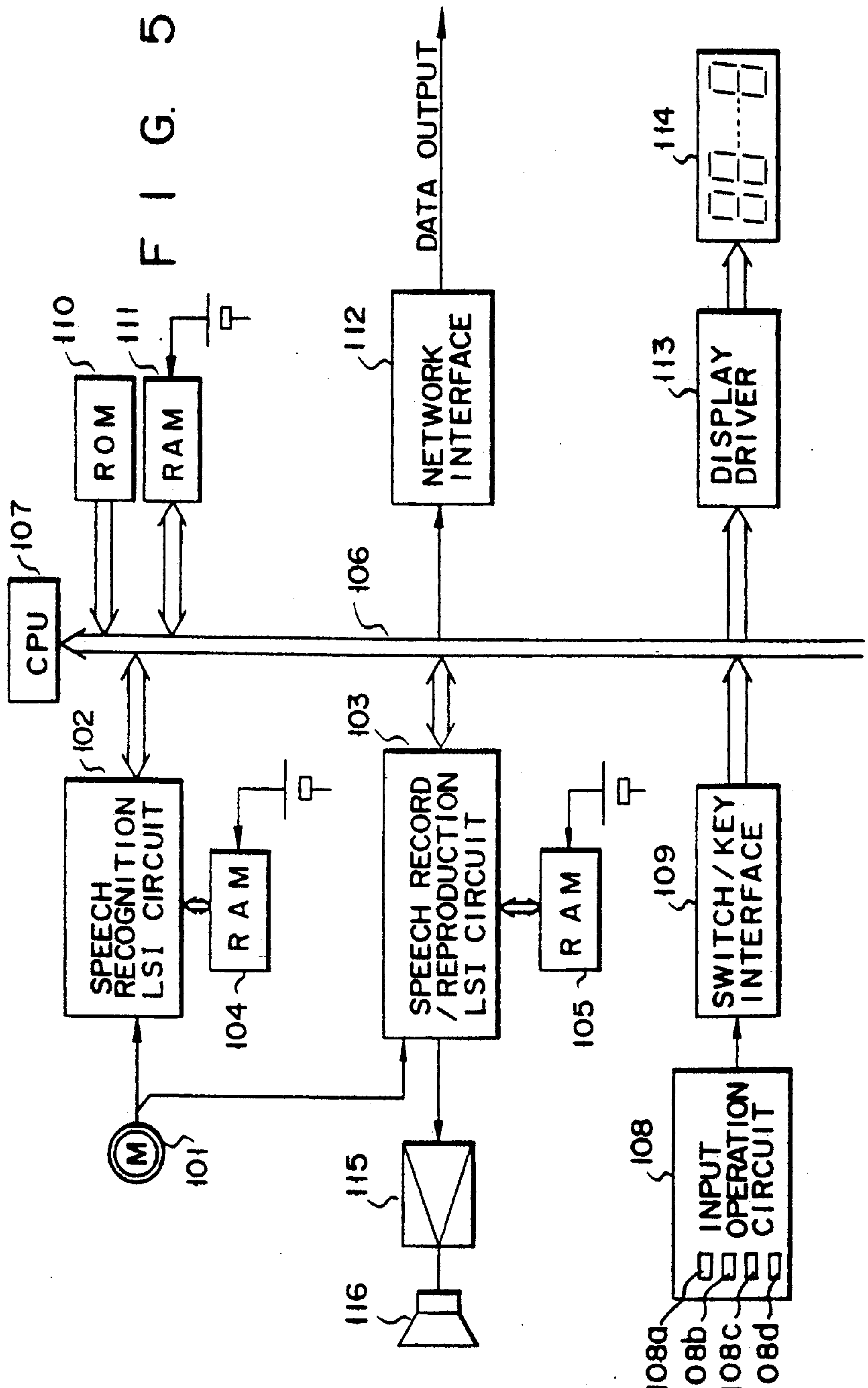

START
4a
RECEPTION ?
YES
4c
CONTROL RECEPTION COUPLING
NO
4b
NO
TRANSMISSION ?
YES
4d
NO
SPEECH INPUT ?
YES
4e
SPEECH RECOGNITION
4f
DIAL NUMBER INFORMATION ?
NO
YES
4g
CONVERT IT TO DIAL NUMBER DATA
B
4h
OUTPUT DIAL NUMBER DATA TO DISPLAY
4i
SPEECH SYNTHESIS
A

FIG. 4A

FIG. 4B

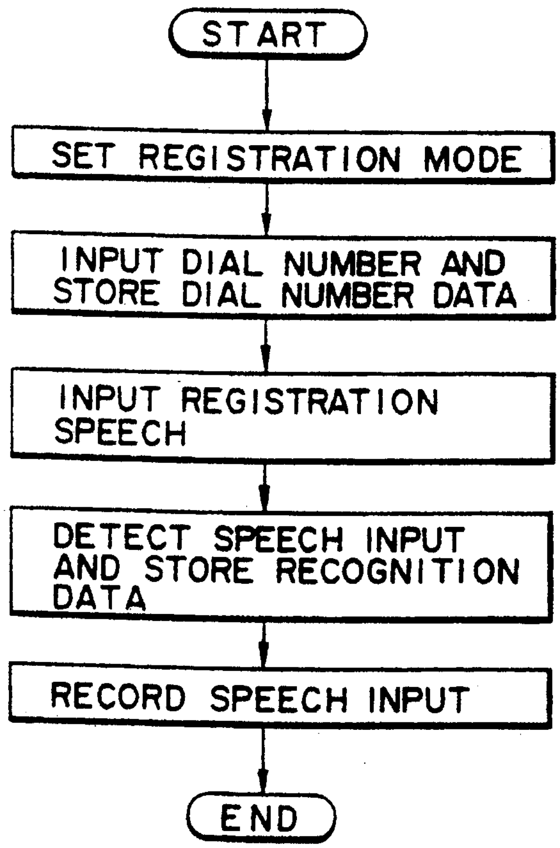

A
OUTPUT SYNTHETIC SPEECH
4j
CONFIRMATION KEY OPERATED ?
4k
NO
B
YES
OUTPUT DIAL SIGNAL
4ℓ
CONTROL TRANSMISSION COUPLING
4m
SPEECH FINISHED ?
4n
NO
YES
END



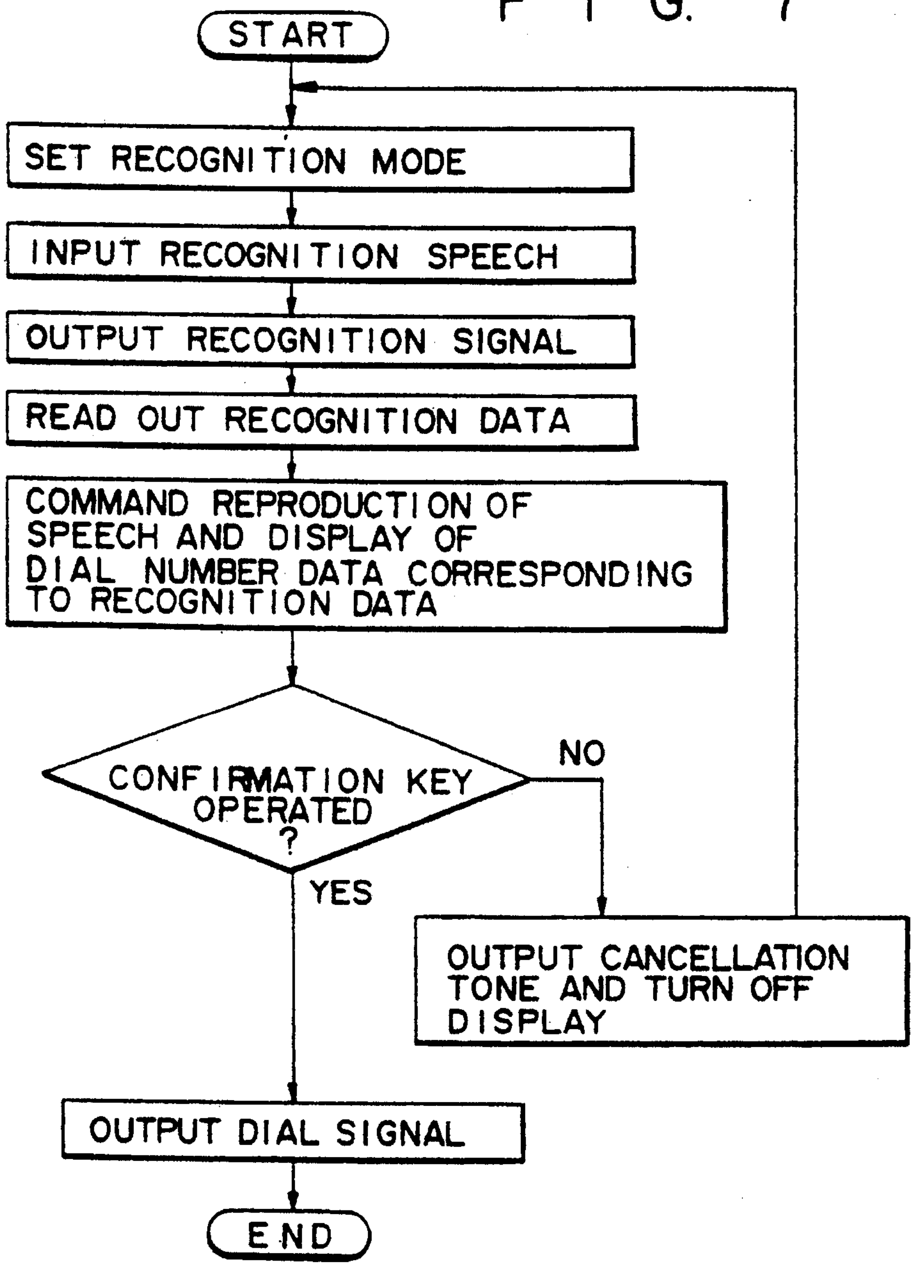
CPU
107
ROM
110
RAM
111
NETWORK INTERFACE
112
DATA OUTPUT
DISPLAY DRIVER
113
114
106
SPEECH RECOGNITION LSI CIRCUIT
102
RAM
104
SPEECH RECORD / REPRODUCTION LSI CIRCUIT
103
RAM
105
SWITCH / KEY INTERFACE
109
M
101
115
116
INPUT OPERATION CIRCUIT
108
108a
108b
108c
108d

FIG. 5

FIG. 6


START
SET REGISTRATION MODE
INPUT DIAL NUMBER AND STORE DIAL NUMBER DATA
INPUT REGISTRATION SPEECH
DETECT SPEECH INPUT AND STORE RECOGNITION DATA
RECORD SPEECH INPUT
END

F I G. 7


START
SET RECOGNITION MODE
INPUT RECOGNITION SPEECH
OUTPUT RECOGNITION SIGNAL
READ OUT RECOGNITION DATA
COMMAND REPRODUCTION OF SPEECH AND DISPLAY OF DIAL NUMBER DATA CORRESPONDING TO RECOGNITION DATA
CONFIRMATION KEY OPERATED ?
NO
YES
OUTPUT CANCELLATION TONE AND TURN OFF DISPLAY
OUTPUT DIAL SIGNAL
END

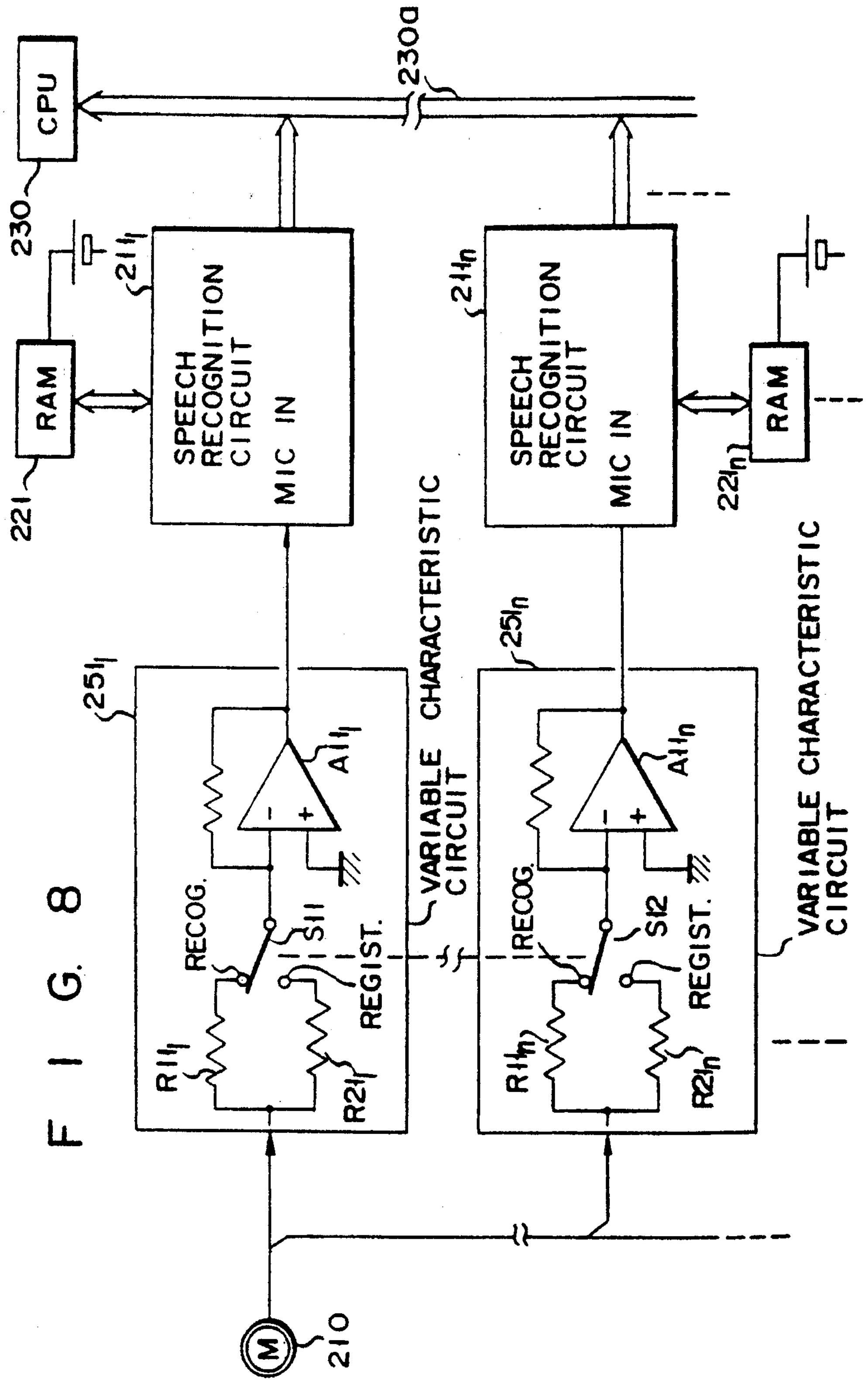

CPU
230
230a
RAM
221
211_1
SPEECH RECOGNITION CIRCUIT
MIC IN
211_n
221_n
251_1
251_n
$A11_1$
$A11_n$
RECOG.
REGIST.
S11
S12
$R11_1$
$R21_1$
$R11_n$
$R21_n$
VARIABLE CHARACTERISTIC CIRCUIT
M
210

FIG. 8

SPEECH RECOGNITION SYSTEM WITH AN ACCURATE RECOGNITION FUNCTION

This application is a continuation of U.S. patent application Ser. No. 07/601,013 filed on Oct. 24, 1990 now abandoned, which was a continuation of U.S. patent application Ser. No. 07/382,346 filed on jul. 20, 1989, now abandoned, which was a continuation of U.S. patent application Ser. No. 06/933,213 filed on Nov. 21, 1986 and which has issued as U.S. Pat. No. 4,873,714.

BACKGROUND OF THE INVENTION

This invention relates to a speech recognition system with an accurate recognition function, and in particular, to a system as one kind of data input device with an accurate recognition function which can eliminate any recognition error.

A conventional speech recognition system of this type comprises a microphone for inputting speech, a speech recognition circuit for registering and recognizing the speech input signal from the microphone, a central processing unit (CPU) for reading a recognition output from the speech recognition circuit and allowing data corresponding to the recognition data to be transmitted to a speech synthesizer via a data bus, a speech synthesizer for receiving data from the CPU and synthesizing sizing the speech from the data, and a speaker for outputting the speech which is synthesized by the speech synthesizer.

According to this system, the user can input data and give an instruction without using his hands, and in fact, it proves very useful when applied to an input device. In this system, however, whether or not the recognition output from the speech recognition circuit is wrong is ascertained by the output of the speech synthesizer. Only predetermined speech patterns are written with respect to the speech synthesizer, placing some restriction upon the speech inputs through the speaker. Since the input speech does not always coincide with the registered speech patterns, it is difficult for the user to ascertain their coincidence. This causes uneasiness in the user. During the speech registration process, it is necessary to register the speech input at a predetermined address position in a memory of the speech recognition circuit, thus involving a cumbersome operation.

Furthermore, since with respect to a respective word or clause one kind of speech pattern is registered for comparison with the speech input, even if the same user inputs his own speech, it is not often successfully recognized due to ambient noise and the delicately varying speech input characteristic. Where, in particular, the speech recognition system is employed as an input device for a mobile station such as in a moving vehicle, the speech is liable to be varied due to the acoustic circumstances within the narrow confines of the vehicle compartment and the traffic noise on the road, resulting in a poor recognition percentage.

Where the speech recognition system is used as the input device for automobile telephones, the telephone set per se never has a speech recognition ascertaining function and, therefore, it is not possible for the user to ascertain that the speech input has been correctly recognized, unless it is sent back from a central station.

SUMMARY OF THE INVENTION

It is accordingly an object of this invention to provide a new and improved speech recognition system with an accurate recognition function, which can momentarily ascertain speech input on the basis of speech patterns initially registered at the time of registration and can register speech input in a memory at any desired address position as well as providing auditory confirmation of the recognition, thus assuring ease of operation by the user.

According to this invention a telephone apparatus as provided in which first subscribers names and corresponding phone numbers were keyed in. Subsequently a user is requested to train the system to recognize the entered names. For each utterance the speech recognition unit generates user-specific templates which can then be referred to during voice dialing operation.

BRIEF DESCRIPTION OF THE DRAWINGS

These and other objects and features of this invention can be understood through the following embodiments by reference to the accompanying drawings in which:

FIG. 1 is a block diagram showing a speech recognition system, according to a first embodiment of this invention, as applied to a wireless telephone apparatus;

FIG. 2 is a block diagram showing the arrangement of a controller in a transmit/receive unit;

FIG. 3 is a block diagram showing the functional arrangement of a CPU in the controller of FIG. 2;

FIGS. 4A and 4B illustrate flow charts showing the control procedure and contents of the CPU in FIG. 3;

FIG. 5 is a circuit diagram showing the arrangement of a speech recognition system, according to a second embodiment of this invention, as applied to a telephone apparatus;

FIGS. 6 and 7 illustrate flow charts showing the operation of the second embodiment; and

FIG. 8 is a block diagram showing a speech recognition system according to a third embodiment of this invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

FIG. 1 shows a circuit arrangement of a speech recognition system according to one embodiment of this invention as being applied to a wireless telephone apparatus. The speech recognition system comprises transmit/receive unit 1, control unit 2 connected to transmit/receive unit 1 through connection cord 4, and headphone type hands-free unit 3.

Transmit/receive unit 1 receives a speech signal sent from control unit 2 through connection cord 4 and connector 11. The speech signal is supplied through amplifier 12 to transmit circuit 13 where it is modulated. The modulated signal is transmitted from antenna 15 to a communication line after it has been passed through common circuit 14.

A signal sent from a telephone apparatus by a called subscriber is received by receive circuit 16 through antenna 15 and common circuit 14. The received signal is, after being amplified by amplifier 17, sent to control unit 2 through connector 11 and connection cord 4. The transmit/receive unit includes synthesizer circuit 18 for designating the transmit and receive channels of transmit and receive circuits 13 and 16, as well as controller 19 as set out below.

Control unit 2 includes microphone 23 as a part of the transmitter and speaker 26 as a part of the receiver. A transmit signal is input through microphone 23 and, after being amplified by amplifier 24, it is sent through connector 21 and connection cord 4 to transmit/receive unit 1. A receive signal, on the other hand, is sent from transmit/receive unit 1 through connection cord 4 and connector 21 to amplifier 25 where it is amplified. The amplified output is delivered to speaker 26 where audible as speech. Control unit 2 includes display 27 made of a liquid crystal, key switch 28 and controller 29. These circuit elements allow a telephone number, input at the time of making a call, to be displayed. Connector 22 links hands-free unit 3 to control unit 2.

Controller 19 in transmit/receive unit 1 is configured as shown in FIG. 2. That is, controller 19 includes main control section (CPU) 90, comprised of a microprocessor and connected via bus 97 to program ROM 91, RAM 92 for data storage, speech recognition circuit 93 and speech synthesizer circuit (speech synthesis circuit) 94. Controller 19 further includes input/output (I/O) circuit 95 for data transfer to and from control unit 2 as well as input/output (I/O) circuit 96 for data transfer to and from synthesizer circuit 18, transmit circuit 13 and receive circuit 16.

As shown in FIG. 3, CPU 90 includes, as main functions, transmission/reception controller 90*a* for controlling a transmit/receive operation, dial number data input controller 90*b* and dial number display controller 90*c*. Dial number data input controller 90*b* permits a call from the user to be input from microphone 31 in hands-free unit 3 and permits dial number speech data, sent through control unit 2 and connection cord 4, to be received via transmit circuit 13, so that the dial number speech input data is recognized by speech recognition circuit 93 to generate dial number code data representing the speech data. Dial number display controller 90*c* supplies code data of the dial number which has been obtained by dial number data input controller 90*b* to speech synthesizer circuit 94 where the dial number is converted to speech data. This dial number speech data is sent through control unit 2 to hands-free unit 3 where it is sounded, as an output, from speaker 32. Dial number display controller 90*c* supplies the code data of the aforementioned dial number to display 27 in control unit 27.

The operation of the aforementioned apparatus will be explained below in accordance with the control procedure of CPU 90 in controller 19.

Prior to using the apparatus, a subscriber registers his own speech patterns in speech recognition circuit 93 by initially inputting his own voice, saying the dial numbers.

In the "wait" state, CPU 90 repeatedly monitors incoming and outgoing signals at steps 4a and 4b as shown in FIG. 4A. If in this state the subscriber issues a calling instruction through the operation of, for example, any keys on key switch 28, then the instruction data is sent from controller 29 in control unit 2 to controller 19 in transmit/receive unit 1 through connection cord 4. By so doing, CPU 90 in control circuit 19 detects the issuance of the calling instruction at step 4b and the process goes to step 4d to await voice input entry of the dial number speech data. When the subscriber inputs the dial number with his own voice through microphone 31 in hands-free unit 3, CPU 90 permits the dial number speech data to be input, digit by digit, from transmit circuit 13 through input/output circuit 96 to speech recognition circuit 93 for speech recognition at step 4e. When voice input which represents, for example, the dial number "03-123-4567," the data "zero", "three", "one" . . . are entered, digit by digit, in that order to speech recognition circuit 93 where the data is compared to the initially-entered speech patterns to allow the data of the corresponding speech pattern to be put to an associated circuit. Upon the completion of speech recognition with respect to the entire digit patterns, the process goes to step 4f at which judgment is made as to whether or not there exists dial number information involved. When there is no corresponding dial number or a wrong dial number, the process goes back to step 4d at which the monitoring operation is performed for speech input. When an incoming signal appears in the aforementioned wait time, CPU 90 goes to step 4c where a reception coupling control operation is performed in a predetermined procedure to establish a conversation circuit.

Where the dial number is recognized, CPU 90 converts the dial number speech pattern data to character code data at step 4g. At step 4h, the data is sent to controller 29 in control unit 2. At step 4l, the character code data of the aforementioned dial number is fed to speech synthesizer circuit 94 to generate corresponding speech. As shown in FIG. 4B, at step 4j the speech data is sent through control unit 2 to hands-free unit 3 and output from speaker 32 in hands-free unit 3. The input dial number is acknowledged both through the visual display by display 27 and through the audible signal from speaker 32 in hands-free unit 3.

When the subscriber confirms the dial number on the display and operates a confirm key on key switch 28, CPU 90 detects the operation data of the confirm key at step 4k. At step 4l, a dial signal is delivered from transmit circuit 13 on the basis of the character code data of the aforementioned dial number. At step 4m a transmit/receive control operation necessary for connection to the called subscriber is performed in the predetermined procedure to establish a conversation circuit at the completion of that control operation, noting that, during a talking phase, CPU 90 monitors speech termination at step 4n and the process goes back to steps 4a and 4b at the termination of conversation.

In the embodiment of this invention, when a dial number is entered at the time of making a call, the dial number is converted to speech data in speech synthesizing circuit 94 in the subscriber's telephone apparatus so that it is produced from speaker 32 in hands-free unit 3. The subscriber, when hearing the speech from speaker 32, can confirm the "now entered" dial number, unless it is sent back from a central station as a mobile network. Even if, therefore, a call is made from a mobile station, such as a moving vehicle, the caller can confirm the dial number without having to look at display 27. The subscriber can perform a dialing operation without endangering his safety. Under any circumstances, it is possible for the subscriber to confirm the dial number and thus largely reduce the occurrence of a "wrong number". In other words, a correct call can normally be made to the called subscriber. Furthermore, according to this embodiment, the dialed number can be confirmed to the calling subscriber by means of both an audible signal and a visual display and, even if the subscriber fails to hear the audible dial number acknowledgement, he can still visually confirm the number on display 27.

Although in the aforementioned embodiment the entry of the dial number data has been explained as being made by spoken word, the same data can also be input through the key switch as required.

Although in the aforementioned embodiment the entry of the dial number data and display of speech data have been explained as being made by microphone **31** and speaker **32** in headphone type hands-free unit **3**, these may be done through microphone **23** and speaker **26** in control unit **2**. Furthermore, the entry of the dial number data may be made by inputting the dial number in an uncurtailed format or by inputting the dial number, if necessary, with the name of the called subscriber, in a curtailed format, in which case CPU **90** determines how the input signal can be converted to the corresponding dial number. The speech recognition and synthesis etc., of the dial number may be made by virtue of controller **19** in the transmit/receive unit or controller **29** in control unit **2**.

Although, in the aforementioned embodiment, subsequent to the entry of the dial number and its judgment for correctness, the conversion to speech data is implemented to allow the whole dial number to be output from speaker **32** in hands-free unit **3**, the dial number may be converted to corresponding speech data for each individual digit entry so that it can be output from speaker **32**. The control procedure and contents of the CPU, configuration and type of the telephone apparatus, configuration of the dial number speech display means, and so on may be properly combined in a variety of ways.

A speech recognition system as applied to a wireless telephone apparatus will now be explained below in connection with second and third embodiments of this invention.

FIG. **5** shows a speech recognition system according to a second embodiment of this invention which can positively ascertain whether or not speech data recognized is correct, without causing any uneasy or uncomfortable feeling in subscribers, a situation which has been encountered in the conventional apparatus.

In FIG. **5** microphone **101** is connected to specified subscriber's speech recognition LSI circuit **102** and speech record/reproduction LSI circuit **103**. RAM's **104** and **105** for data storage are connected to LSI circuits **102** and **103**, respectively, which in turn are connected via data bus **106** to processing circuit (CPU) **107**. CPU **107** is adapted to perform predetermined control processings necessary for telephone apparatuses, such as the registering, recognition and reproduction (acknowledge). Input operation circuit **108** is connected via switch/key interface **109** to data bus **106** and includes mode select switch **108a**, ten keys **108b** for telephone number registry, confirm key **108c** and negation key **108d**. To data bus **106** are connected ROM **110** for storing programs necessary for the aforementioned control processings, RAM **111** for telephone number storage and communication line interface **112**. Display **114** is connected via display driver **113** to data bus **106**. Speaker **116** for producing a reproduction speech is connected through amplifier **115** to speech record/reproduction LSI circuit **103**.

The LSI circuits **102** and **103** are conventional and may be made with various LSI's which are readily available, for example, T6658a (Toshiba) for circuit **102** and TC 8830 (Toshiba) for circuit **103**. For the internal configuration and detailed connections reference is invited to the technical data attached to T6658A and TC8830, noting that the same thing is also true of a third embodiment as set out below.

The operation of the aforementioned embodiment will be explained below.

At the time of registry, the user first sets this apparatus to a speech registration mode through the operation of input operation circuit **108**. Then he inputs the telephone number data of the called subscriber through the operation of ten keys on input operation circuit **108** at the time of making a call. The telephone number data thus input is stored in RAM **111** under the control of CPU **107**. Then the user inputs the registered speech, such as the name of the called subscriber, by speaking into microphone **101**. The voice input of the user is detected by detecting that the voltage level of the input signal exceeds a predetermined threshold level. The detected voice input is stored as recognition data in RAM **104** under the control of CPU **107**. Concurrently the utterance input is delivered as record data from microphone **101** through speech record/recognition LSI circuit **103** to RAM **105**, under the control of CPU **107**, where it is stored. The aforementioned register processing is repeated a necessary number of times. A tape recorder may be employed in place of speech record/reproduction LSI circuit **103**.

After the apparatus has been set to the speech recognition mode through the operation of the switch on input operation circuit **108**, the user speaks a corresponding recognition word into microphone **101** in the same way as set forth above. Upon receipt of the output from microphone **101**, speech recognition LSI circuit **102** delivers a recognition signal, as an interrupt signal, to CPU **107** in which case the output from the microphone **101** is not coupled to speech record/reproduction LSI circuit **103**. CPU **107** reads the recognition data corresponding to the recognition speech from RAM **104** through speech recognition LSI circuit **102**. CPU **107** issues an instruction for reading the record data corresponding to the read-out recognition data from RAM **105** to permit reproduction, as well as an instruction for reading the called subscriber's telephone number data corresponding to the recognition data from RAM **111** and supplying it to display driver **113** for display. By so doing, the user can judge whether or not the recognition data is correct on the basis of the reproduction speech from speaker **116**, without causing any uneasiness in the user, noting that the reproduced speech is not a restricted synthesized one as in the conventional apparatus, but his own speech, which is of importance according to this invention. In addition to this judgment the user can also ascertain it, as required, on display **114**.

Where the user recognizes the aforementioned judgment as being correct, CPU **107** allows the called subscriber's telephone number data corresponding to the recognition data to be read out of RAM **111** by a "confirmed" response, or a lack of response over a predetermined period of time, from confirm key **108c** on input operation circuit **108**, except in the case of a mere initial acknowledgement, so that it may be fed to network interface **112**.

Where the user recognizes the aforementioned judgment as being wrong, CPU **107** delivers a tone or speech data representing a negation or cancellation response made through the operation of negation key **108d** on input operation circuit **108** and, at the same time, allows the initial state to be regained, in place of the speech recognition mode, with display **114** extin-

guished. In this case, a registration is made, as required, in the aforementioned registration mode.

FIGS. 6 and 7 show flow charts corresponding to registration and recognition modes, respectively.

The user can register his own speech data in any desired address position through the registration/reproduction LSI circuit, thus assuring a ready system operation.

As set out above, according to this invention the speech contents upon registry can be made coincident with the speech contents upon recognition and at the same time the speech input can be registered by the user in any desired address position, whereby a ready availability and operability are assured.

Since, in particular, a call is made over the telephone through the verbalization of the subscriber's name, the incidence of "wrong number" can positively be prevented according to this invention.

FIG. 8 shows, as a third embodiment of this invention, the major section of a speech recognition system having the function of positively recognizing spoken word when it is affected by the manner of speaking as well as by the circumstances under which it is spoken.

The speech recognition system includes a plurality of speech recognition circuits **211**$_1$, . . . **211**$_n$ and a corresponding number of variable characteristic circuits **251**$_1$, . . . **251**$_n$. Speech recognition circuits **211**$_1$, . . . **211**$_n$ are connected to speech registration RAM's **221**$_1$, . . . **221**$_n$, where a plurality of modified speech patterns are stored for each word or each clause. The modified speech patterns include, as the characteristics of the speech, the magnitudes of sound volumes, the varying lengths of sounds, the variation of sound intervals, the husky or somewhat masked voices, and so on. The variable characteristic circuit (**251**$_1$. . . or **251**$_n$ is comprised of an amplifier (**A11**$_1$, . . . or **A11**$_n$) of a different amplification factor and different bandpath characteristic, set of resistors (**R11** and **R21**$_1$. . . or **R11**$_n$ and **R21**$_n$; of 2:1 values and mode designating switch (**S11**$_1$, . . . or **S11**$_n$) to allow the speech which is input through microphone **210** in a speech registration mode to be prepared as a variation pattern so that it is stored as such in speech recognition circuit (**211**$_1$. . . or **211**$_n$).

When the speech is to be registered, variable characteristic circuits **251**$_1$. . . **251**$_n$ and speech recognition circuits **211**$_1$, ... **211**$_n$ are set to speech registration modes with mode designation switches **S11**$_1$, . . . **S11**$_n$ set to the speech registration mode. In this state, if the user inputs a desired word through microphone **210** by saying the object word to be recognized, then the speech signal is branched to variable characteristic circuits **251**$_1$. . . **251**$_n$ where the branched speeches are modified in accordance with their amplification factor and bandpath characteristic. If, for example, the word "one" is input through the microphone, then the sound characteristics, such as the volume, length, interval and quality for "one", are subjected to modification processing. That is, the variable characteristic circuits **251**$_1$, . . . **251**$_n$ prepare n kinds of speech patterns which are obtained through the arbitrary modification of the sound characteristics for "one." The respective modification speech signal is supplied to speech recognition circuits **211**$_1$. . . **211**$_n$ so that they may be stored in speech registration RAM's **221**$_1$. . . **221**$_n$. In this way, the speech patterns are registered for one word. The other words and clauses are so registered in exactly the same way.

When a speech registration is to be made in this system, the mode designation switches **S11**$_1$, . . . **S11**$_n$ are set to the speech registration mode REGIST.

When the user inputs his speech through microphone **210** it is, after being branched, supplied respectively through variable characteristic circuits **251**$_1$, . . . **251**$_n$ to speech recognition circuits **211**$_1$, . . . **211**$_n$ where the aforementioned input speeches are compared to the speech patterns which have been registered in the respective RAM's **221**$_1$. . . **221**$_n$. The results of comparison are fed via data bus **230a** to CPU **230** where the speech recognition processing is performed based on the results of comparison. If, for example, the input speech coincides with any one of the registered speech patterns as the result of comparison, the input speech is recognized from the coincidence data.

According to this embodiment, since the input speech is compared to the plurality of modified speech patterns through speech recognition circuits **211**$_1$, . . . **211**$_n$, the speech input can be recognized with a high probability even if it is somewhat modified due to, for example, the illness of the user or the ambient noise.

Even if the user employs this system on a mobile station such as an automobile where the quality etc., of his speech is adversely affected, the speech input can be recognized with a high recognition percentage in spite of the degraded acoustic environments and noise in the moving vehicle.

Through variable characteristic circuits **251**$_1$, . . . **251**$_n$ corresponding to speech recognition circuits **211**$_1$, . . . **211**$_n$ the user has only to input a word or clause, one unit after another, by speaking, as in the case of the conventional system, making it easier for the user to perform a speech registration operation.

This invention is not restricted to the aforementioned embodiments. Although a plurality of modified speech patterns are prepared for registration through the utilization of variable characteristic circuits **251**$_1$, . . . **251**$_n$ for speech recognition circuits **211**$_1$, . . . **211**$_n$, the user can input modified speeches directly through microphone **210** without utilizing the variable characteristic circuits.

With respect to the number and kinds of modified speeches, their registering means, their speech recognizing means, etc., can be changed or modified within the spirit and scope of this invention.

What is claimed is:

1. A speech recognition system comprising:

microphone means for producing speech input signals of various users, to be registered or recognized;

speech recognition and registration means responsive to a speech input signal from the microphone means, so as to subject the speech input signal to either a registration or a recognition processing, in which upon the registration processing, the speech input signal is allowed to be stored as recognition data and upon the recognition processing, the speech input signal is compared to the recognition data which has been stored; and

a central processing unit which performs speech recognition processing based on the results of comparison; wherein said speech recognition and registration means comprises a plurality of speech recognition circuits which are connected between a corresponding number of variable characteristic circuits and speech registration RAM's respectively where a plurality of modified speech patterns are stored for each word or each clause, said modified speech

patterns including, as the characteristics of the speech, the magnitudes of sound volumes, the varying lengths of sounds, the variation of sound intervals, the husky or somewhat masked voices, and

each of said variable characteristic circuits is comprised of an amplifier of a different amplification factor and different bandpath characteristic, a set of resistors and a mode designating switch, respectively, to allow the speech which is input through said microphone means in a speech registration mode to be prepared as a variation pattern so that it is stored as such in the speech registration RAM of the corresponding speech recognition circuit, such that when a user inputs his speech through said microphone means it is supplied respectively through said variable characteristic circuits to said speech recognition circuits where the input speeches are compared to the speech patterns which have been registered in the respective speech registration RAMs.

* * * * *

EXHIBIT 9

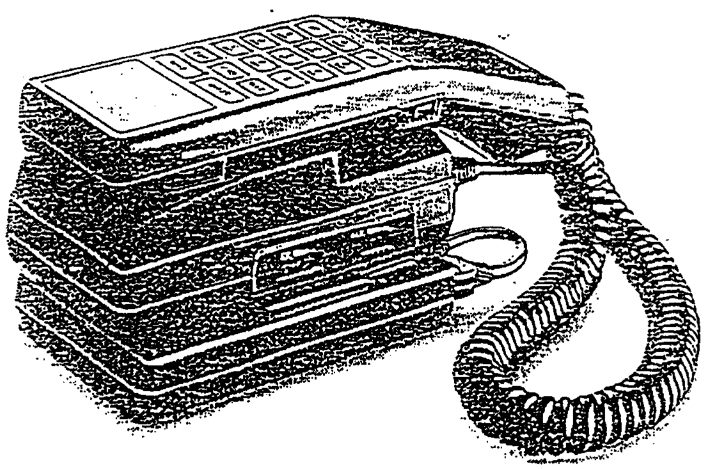
C

VoiceDial

OPERATING GUIDE

America's First Speaker Independent

Voice Command System For
Cellular Telephones



uniden®

Table Of Contents

System Technologies supplied by
Voice Control Systems and By-Word Technologies, Inc.

INTRODUCTION

Welcome to Uniden America Corporation's new world of voice command phone operation! Using your cellular phone with **VoiceDial** is not only easy, it adds safety, speed and convenience. **VoiceDial** permits normal phone functions, besides adding voice command features as well. Now, keep both hands on the steering wheel and your eyes on the road while you make or receive your telephone calls.

Before you begin, please take a few minutes to read through this Guide. It contains all the information you need to get started, plus useful hints to help you get the best performance from **VoiceDial**.

KEEP THIS GUIDE IN THE GLOVE COMPARTMENT OF YOUR VEHICLE FOR REFERENCE.

VOICE COMMAND YOUR PHONE

You can voice command your phone in a number of ways:

- Dictate the digits of the phone number you want to dial.
- Say one of ten descriptive words to dial preprogrammed numbers. For example, *"Office"* to dial your office number, *"Home"* to dial your home number, etc.
- Access 19 stored numbers from your phone's memory and dial this number with a key word.
- Hang up your phone by a simple command.
- Answer incoming calls by voice command.

We hope that you use all of these features when you make and receive your phone calls, but whichever feature's you use, **VoiceDial** puts you in command of your phone!

STARTING VOICEDIAL

◊ Power On - When you turn the power on to your phone, you hear several tones. This indicates VoiceDial is testing itself and when completed, you hear *"Hello."* This occurs each time you turn the ignition on.

◊ Standby - After the ignition is turned on and VoiceDial has completed its testing, VoiceDial is in standby and waits until you say the key words to start.

◊ Key words - To start VoiceDial, say the two key words, *"Phone"*...*"Start"* with a one second pause between them.

NOTE: Be sure to leave a one second pause between the two words. This pause prevents Voicedial *from accidentally starting from conversations inside your vehicle.*

If too short of a pause is left between the words, VoiceDial may: Remain silent while waiting for the second word, say *"Slower please"* which indicates not a long enough pause between the words, or other responses.

We recommend that you take a few minutes to practice this feature. (Listen to the provided audio cassette tape to hear how VoiceDial works.)

◊ Practice Key Words - Turn the power on and press * * 7 on your phone keypad. (This code is only used for the practice session.)

Say the words *"Phone"*...*"Start,"* in a normal voice while facing the microphone (located 12 to 18 inches away). If VoiceDial understands each of the three elements, you hear *"Ready"* and on the phone display you see three 1's for about one second. Each digit 1 indicates a correct recognition for each element.

If any one of the three elements (first word, a pause, second word) is not correct, VoiceDial says *"Error"* and the display shows a digit 0 for each incorrect element.

To end the practice session, press any key, otherwise it is automatically deactivated after a one and a half minute time out which returns VoiceDial to standby.

Should you still have trouble with starting VoiceDial, call Uniden technical support at (317) 842-2483. The practice session also works during a call-in-progress. You are first asked if you heard *"Hello"* from VoiceDial. If not, press the buttons * * 1 and then press * * 7 and say the key words *"Phone"*. . .*"Start."* Now you are asked what appeared on the phone display and they can tell you where the trouble is. During your call, * * 7 must be entered to see the recognition for each element result.

NOTE: If for any reason you cannot start Voicedial *by voice command, press the* END *button.*

◊ VoiceDial responds with *"Ready"* and a beep.

The response *"Ready"* means that VoiceDial is waiting for you to speak a command. If you do not say a command in a short period of time, VoiceDial puts itself in standby.

As soon as you hear *"Ready,"* make your phone call with VoiceDial in one of four ways:

DIAL BY DICTATING DIGITS

◊ Say *"Phone"*...*"Start."* You hear *"Ready,"* and a beep.

◊ Say *"Dial."*

◊ VoiceDial responds with *"Number please."*

◊ Say the phone number, speaking one digit at a time, say the first digit and pause until you hear a beep, then proceed to the next digit (each digit is shown on the phone display as you speak it). Wait until you hear this beep before you say the next digit.

◊ After speaking the last digit and hearing a beep you must say *"End."* This tells VoiceDial that you have finished speaking the digits.

◊ VoiceDial repeats the number back to you and also shows the telephone number on the phone display.

◊ Say *"Send."*

◊ VoiceDial says *"Dialing,"* and dials the number.

At any time during this sequence, if you or VoiceDial make a mistake, simply say *"Clear."* The last digit is cleared off the phone display and VoiceDial says *"Digit please."* You then enter a new digit.

To clear the whole display, say another *"Clear."* To dictate the number again, begin with *"Dial"* and then the digits as outlined above.

On occasion, VoiceDial does not understand what you have said and asks you to *"Repeat please"* or *"Pardon."* If this happens, repeat the last digit or last command only. Do not repeat the whole sequence of digits.

VoiceDial may also say *"Louder please."* If this happens just follow the request and speak the command accordingly.

The response *"Slower please"* means VoiceDial needs a longer pause between commands.

REDIAL THE LAST NUMBER DIALED

Use this feature to redial phone numbers that you could not get through (the line is busy or there is no answer).

To use the Last Number Dialed feature, do the following:

◊ Say *"Phone"*...*"Start."* You hear *"Ready,"* and a beep.

◊ Say *"Send."* Voice Dial calls the last number dialed.

NOTE: This number may not appear on the phone display.

If you do not remember what the last number you dialed was, do the following:

◊ Say *"Phone"*...*"Start."* You hear *"Ready,"* and a beep.

◊ Say *"Call."*

◊ VoiceDial says *"Calling."*

◊ Say *"Memory."*

◊ VoiceDial asks *"Which memory?"*

◊ Say *"Zero"*...*"Zero"* Wait for the beep after each digit.

Say *"Send"* and VoiceDial dials the call, or say *"Verify"* and the phone number is repeated back to you before you say *"Send."* This is convenient to check the last number dialed contents before making the call.

At any time you can say *"Clear"* and VoiceDial goes back to *"Ready."* To completely abandon the dialing sequence say *"Clear"* again. You hear 2 beeps. VoiceDial is in standby.

HANG UP THE PHONE

After you finish your conversation and want VoiceDial to hang up the phone:

◊ Say *"Phone" . . . "Quit."* Pause about one second between each word. (Pressing the END button on your phone also hangs up the phone.)

◊ You hear one beep followed by a pause and then a double beep.

◊ VoiceDial automatically hangs up the phone and is in standby.

NOTE: If you cannot hang up the phone by voice command (this occurs while attempting to make a call and the phone at the other end is still ringing or busy), press the END button.

During a conversation there may be an occasion when you hear a beep. VoiceDial, while in standby, mistakingly heard the command to hang up the phone. If this happens, ignore the beep and continue talking. VoiceDial recognizes you are still talking and keeps the line open.

After your conversation has ended and you want to immediately dial another number, do the following:

◊ Say *"Phone" . . . "Start"* instead of *"Phone" . . . "Quit."*

◊ VoiceDial hangs up the first call, then says *"Ready"* and waits for your voice commands to dial the second call.

ANSWER INCOMING CALLS

When the phone rings, to answer the incoming call with VoiceDial:

◊ Say *"Phone."* You must speak this command between rings. This may take several attempts to accomplish.

◊ VoiceDial says *"OK"* and now you can talk to your caller with the handsfree microphone.

When you have finished your conversation, to hang up the phone with VoiceDial:

◊ Say *Phone"* . . .*"Quit,"* or press the END button.

STORE TELEPHONE NUMBERS BY VOICE

To use VoiceDial's Dial By Descriptive Words or Dial Stored Numbers In Memory, do the following:

To store your home telephone number using the descriptive word *"Home"*:

◊ Say *"Phone"*. . .*"Start."* You hear *"Ready,"* and a beep.

◊ Say *"Dial."*

◊ VoiceDial says *"Number please."*

◊ Say your home number, one digit at a time, waiting for the beep after each digit.

◊ After saying the last digit and hearing the beep, say *"End."*

◊ VoiceDial repeats the number to you.

◊ Say *"Store."*

◊ VoiceDial respon is with *"Storing."*

◊ Say *"Home."* (*See* Store Numbers In Memory *below*).

◊ VoiceDial says *"Storing Home,"* and proceeds to store your home number. When finished storing, VoiceDial says *"Ready."*

◊ You can now store another phone number against another VoiceDial descriptive word using the same procedure as above, beginning at the command *"Dial,"* or finish by saying *"Clear"* or press the END button.

NOTE: See the Store By Voice Procedure *chart on page 9.*

Store Numbers In Memory – Store telephone numbers in the phone's memory locations 01 through 19 as follows:

◊ Instead of saying *"Home"* as in above, say *"Memory."*

◊ VoiceDial asks *"Which memory?"*

◊ Say the two digit memory location 01 - 19, one digit at a time. You must say two digits for a memory location. Example: *"Zero"* . . .*"Two"* for memory location 2, or *"One"* . . .*"Six"* for memory location 16. Wait for the beep after each digit.

◊ VoiceDial says *"Storing memory . . ."* and stores the phone number in the designated memory location.

NOTE: See the Store By Voice Procedure *chart on page 9.*

Keep a record of the numbers you are storing in the memory locations in the back of this manual on page 17. You may need to refer to this list in the future.

NOTE: Any number stored with this procedure can be erased by over writing with a new number.

STORE BY VOICE PROCEDURE

Say *"Phone"*...*"Start"* or press END

"Ready"

Say *"Dial"*

"Number please"

Say each digit with about a one second pause

After the last digit say *"End"*

Number is repeated

Say *"Store"*

"Storing"

Say only one

"Home"
"Office"
"Secretary"
"Information"
"Emergency"
"Service"
"Friend"
"Broker"
"Work"
"Bank"

"Storing..."

Say *"Memory"*

"Which memory?"

Say the two digit memory location 01 - 19

"Storing memory..."

To end the procedure say *"Clear"* or press END

DIAL BY DESCRIPTIVE WORDS

VoiceDial can dial numbers associated with any of the ten descriptive words listed below.

NOTE: Before using this feature, first store telephone numbers to some or all of the descriptive words listed below. Do this by using the procedure in Store Telephone Numbers By Voice.

◊ Say *"Phone"*. . .*"Start."* You hear *"Ready,"* and a beep.

◊ Say *"Call."*

◊ VoiceDial says *"Calling."*

Say one of the words from VoiceDial's descriptive word list:

"Home"	"Service"
"Office"	"Friend"
"Secretary"	"Broker"
"Information"	"Work"
"Emergency"	"Bank"

◊ VoiceDial says *"Calling. . ."* and the descriptive word.

You now have a choice: say *"Send"* and VoiceDial dials the call, or say *"Verify"* and the telephone number is repeated back to you before you say *"Send."*

At any time during this sequence you can say *"Clear"* and VoiceDial goes back to *"Ready."* To completely abandon the dialing sequence say *"Clear"* again. You hear 2 "beeps." VoiceDial is in standby.

NOTE: See Dial/Answer By Voice Procedure *chart on page 12.*

DIAL STORED NUMBERS IN MEMORY

VoiceDial can access the phone's memory to retrieve numbers that you have stored there. Valid memory locations are 01 through 19.

NOTE: Before using this feature, first store telephone numbers to some or all of the memory locations 01 through 19. Do this by using the procedure in Store Telephone Numbers By Voice.

◊ Say *"Phone"*...*"Start."* You hear *"Ready,"* and a beep.

◊ Say *"Call."*

◊ VoiceDial says *"Calling."*

◊ Say *"Memory."*

◊ VoiceDial asks *"Which memory?"*

◊ Say the two digit memory location. You must say two digits for a memory location. Example: *"Zero"* ...*"Two"* for memory location 2, or *"One"* ...*"Six"* for memory location 16. Wait for the beep after each digit.

NOTE: Your Uniden phone, memory locations 20 through 30 are no longer accessible. If you say any one of these locations by mistake, VoiceDial responds with "Location error" and then asks you for a new command by saying "Memory please." Now say a valid memory location 01 and 19.

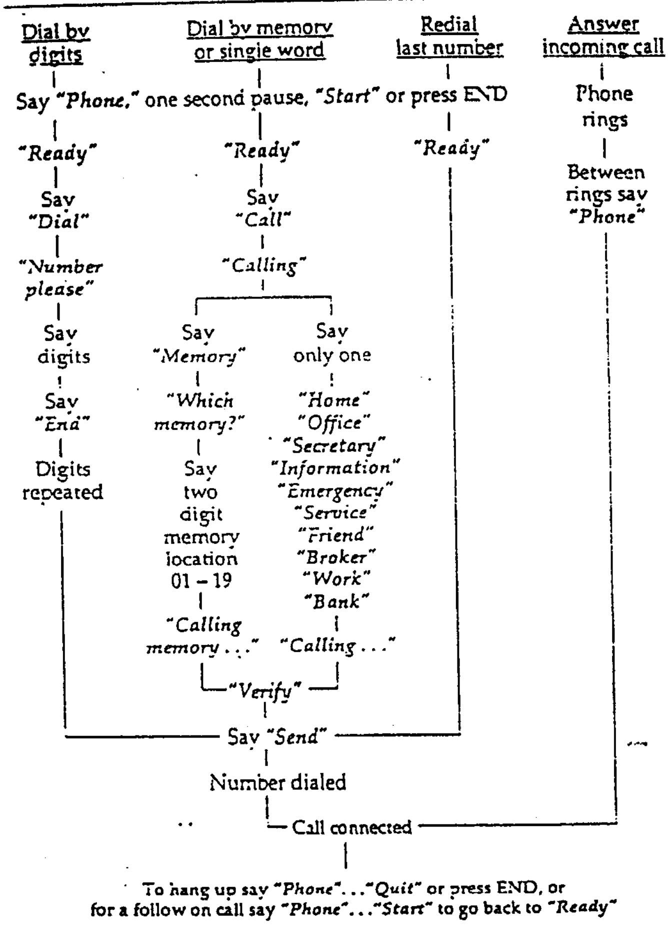
You now have a choice: say *"Send"* and VoiceDial dials the call, or say *"Verify"* and the phone number is repeated back to you before you say *"Send."* This is convenient to check the memory location's contents before making the call.

At any time during this sequence you can say *"Clear"* and VoiceDial goes back to *"Ready."* To completely abandon the dialing sequence say *"Clear"* again. You hear 2 "beeps." VoiceDial is in standby.

NOTE: See Dial/Answer By Voice Procedure *chart on page 12.*

SS 04328

DIAL/ANSWER BY VOICE PROCEDURE



◊ During a sequence you can say *"Clear"* to bring you back to *"Ready."*

◊ For accurate results follow the sequences exactly as shown.

◊ If you get error messages, please refer to VoiceDial Messages And Responses.

USER PROGRAMMABLE FEATURES

Some features of VoiceDial can be turned on and off by pressing certain codes on the keypad of your phone. All the codes act like an on/off toggle switch. VoiceDial is designed with this flexibility for each user's needs.

Each time you change a setting, VoiceDial tells you what its voice response will be. Listed below are all the features you can change. When VoiceDial leaves the factory, all features are set to on except for Digit Repeat.

FEATURE	KEY CODE	VOICE RESPONSE ON	VOICE RESPONSE OFF
DIGIT REPEAT (Each digit repeated as dictated)	* * 0	*"Digit repeat"*	*"Digit repeat off"*
PHONE START (Turned on)	* * 1	*"Hello"*	*"Hello off"*
PHONE QUIT (Hang up)	* * 2	*"Service OK"*	*"Service off"*
PHONE (Call answer)	* * 3	*"OK"*	*"OK off"*
PHONE START PRACTICE (User practice session)	* * 7	—	—

RECOMMENDATION: Enter these codes into your phone only when the phone is powered up and your vehicle is stationary.

NOTE: With some cellular phones, the Digit Repeat *and* Phone Start Codes *are opposite to the above list. If this occurs, simply use the other code for the proper response.*

VOICEDIAL MESSAGES AND RESPONSES

VoiceDial may give you a response for a number of reasons: if you mispronounce or say an incorrect command; noises in the background, such as other people talking, the radio is loud, unusually loud traffic noises, etc. The most important of the responses you might hear are *"Please repeat,"* or *"Pardon," "Slower please"* and *"Louder please."* You may also hear *"Location Error," "Memory Error"* or just *"Error."*

- *"Please repeat"* or *"Pardon,"* means you should repeat the last word or digit that you spoke.
- *"Slower please"* means you should say again, but leave a slightly longer pause between the words or digits.
- *"Louder please"* means you should repeat the last word or digit that you spoke, only this time louder.

Typically, one of these responses has been generated because VoiceDial heard something that it did not expect or cannot understand. Usually, it is obvious what caused the response (such as a loud background noise, etc.).

However, it is not always possible to control such noises in a moving vehicle. If conditions are not right, say *"Clear"* to get back to *"Ready,"* and then another *"Clear"* again, or press END. You hear 2 "beeps." VoiceDial is in standby.

- *"Location error"* occurs if you ask VoiceDial to access an invalid number from the phone's memory. For example, if the phone has twenty available memory locations and you give the commands *"Call"* , *"Memory," "Six"* . . .*"Two"* (i.e. memory location 62), VoiceDial knows there is no such memory location.
- *"Memory error"* occurs if you ask VoiceDial to recover a telephone number from an empty memory location.
- *"Error"* occurs if you dictate more than sixteen digits at any one time.

If you hear any of these error messages, it is best to stop and check this Guide to see where the mistake is being made and start from the beginning again.

In addition to the above messages, there are other responses. These messages tell you about VoiceDial's status or remind you to say something. All these responses are self explanatory and you should have no difficulty in understanding what they mean or why they are prompting you.

HINTS IN OPERATING VOICEDIAL

The following hints are helpful in getting the best performance from VoiceDial:

- When you say *"Dial,"* VoiceDial expects to hear only the digits, followed by the command *"End."*
- When you say *"Call,"* VoiceDial expects to hear one word from the descriptive word list and only these words.
- When you say *"Call"* followed by *"Memory,"* VoiceDial expects to hear a two digit memory location.
- A single *"Clear"* undoes your last command and brings you back to *"Ready"* (unless you are dictating digits, then the last dictated digit is erased from the display and VoiceDial says *"Digit please"* to enable you to say another digit. If you now want to stop dictating digits a second *"Clear"* is required to bring VoiceDial to *"Ready"*).

  A final *"Clear"* after *"Ready"* puts VoiceDial in standby.
- VoiceDial does not interfere with the normal functions of your phone. Make and receive your calls as normal.
- You can always turn VoiceDial on and off by pressing the END button. When in doubt, press the END button to start over. Press buttons lightly – never hold them down.

- While making a call and the phone at the other end is still ringing or busy, you can not hang up the phone by voice command. If either of these conditions occurs, hang up your phone by pressing the END button.
- Shouting distorts your voice. VoiceDial may interpret this as noise. Always speak clearly and in a natural tone. Digits should be dictated one at a time, but wait for the beep between each digit.
- The microphone should be mounted directly in front of you, positioned 12 to 15 inches from your mouth and the switch on LO, or 18 inches on HI. The visor is a good location, since a directional microphone performs best when you face the microphone when you speak.
- For the best performance, try to make your calls when background noises are at a minimum. VoiceDial has been designed to accept a certain amount of noise, but loud noises inside and outside a moving vehicle can overwhelm it. Windows should be closed, the volume on the car radio or stereo turned down, and the heater/air conditioner fan set at an intermediate level. Keeping background noises at a minimum helps the other party hear you more clearly when you talk on the handsfree microphone.
- If other people in your vehicle are talking at the same time, try not to give voice commands to VoiceDial.
- Do not eat or drink when giving voice commands.
- If you receive a call while you are in the middle of voice dialing, VoiceDial immediately abandons its tasks, puts itself in standby and waits. The phone rings normally and you can then answer the call by saying *"Phone."*
- VoiceDial alerts you to a hang up with a beep. Should you hear a beep during your conversation, to avoid a "false" hang up, just continue talking.

STORED NUMBERS DIRECTORY

Write in your frequently called telephone numbers:

Mem. Loc.	Name	Phone No.
01		
02		
03		
04		
05		
06		
07		
08		
09		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
Home		
Office		
Secretary		
Information		
Emergency		
Service		
Friend		
Broker		
Work		
Bank		

EXHIBIT 10

TARGET TOPIC

AUTOMOTIVE

DIALING A PHONE BY VOICE

BASAVARAJ I. PAWATE
PETER EHLIG
Technical Staff Members
Texas Instruments Inc.
Houston, TX

Soon you may be able to "dial" a car phone and turn on the lights and wipers with voice commands.

Look for speech recognition to be the next hot technology in the burgeoning automotive electronics industry. In fact, some experts expect voice command systems that control vehicle functions to become widely accepted in this decade.

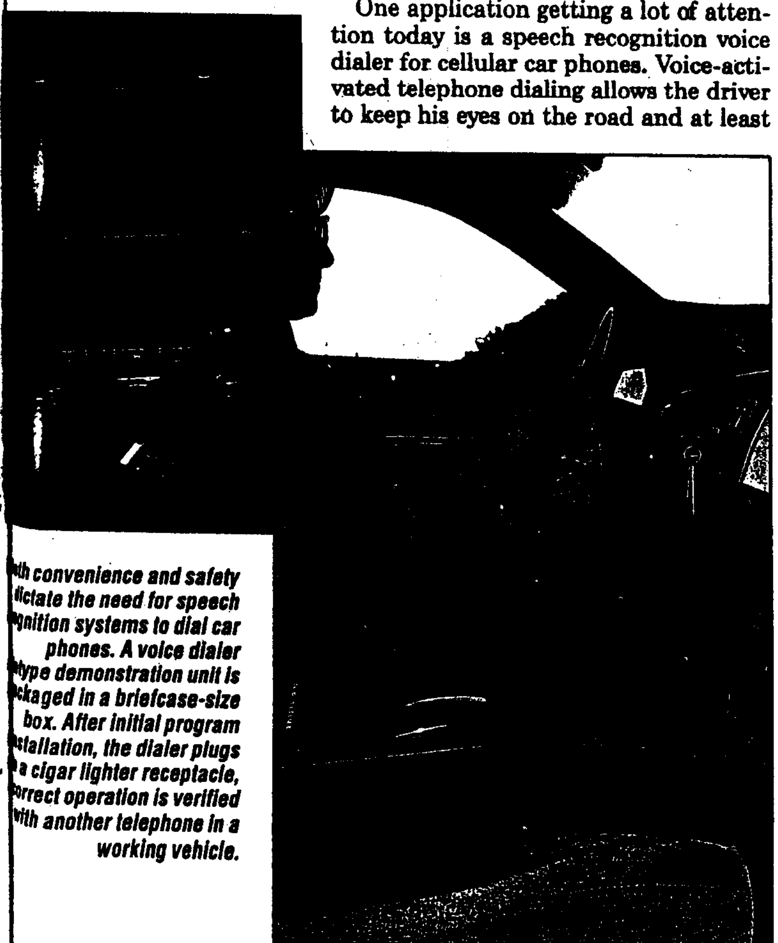
One application getting a lot of attention today is a speech recognition voice dialer for cellular car phones. Voice-activated telephone dialing allows the driver to keep his eyes on the road and at least one hand on the wheel. Conventional dialers, in contrast, require operators to look at a keypad to punch in numbers, a dangerous activity in moving vehicles.

The voice dialer recognizes both male and female voices, as well as a number of dialects. It can have a vocabulary of 25 or more words, depending on memory size. Surprisingly, all this functionality requires only one digital signal processor (DSP).

The voice dialer employs a speech recognition algorithm known as continuous density Hidden Markov Modeling (HMM). HMMs are statistical models for vocabulary words. The algorithms devised to decode voice patterns require substantially more computing power than other techniques, but the improved recognition accuracy outweighs any added expense incurred by using bigger microprocessors.

The voice recognition system has a speaker-independent mode, which means a person does not have to train it to learn his or her voice. For example, any rental-car customer can use the dialer. Any American speakers, regardless of their accents, can be accommodated. Continuous speech recognition is employed so the speaker can talk naturally; no deliberate pauses between words are required.

In addition to unsurpassed accuracy, the voice dialer solves a related communications problem. The cellular telephone industry is rapidly running out of available channels because of the demand for such service. However, a new algorithm called Vector Sum Excited Linear Predictive (VSELP) speech coding, allows the



...th convenience and safety ...dictate the need for speech ...gnition systems to dial car phones. A voice dialer ...type demonstration unit is ...ckaged in a briefcase-size box. After initial program ...stallation, the dialer plugs ...a cigar lighter receptacle, ...orrect operation is verified with another telephone in a working vehicle.

TARGET TOPIC

phone system to accommodate mc channels in the available bandwidth tha.. previous methods.

Using the dialer

A typical application uses a grammar definition program built into, or down-loaded to, the DSP memory, so either man or woman can speak to a car tel- phone and say "Call office" or "Ca: home." He or she can also state the num. ber to be called, using the words zer through nine for digits or the word "oh for zero. The user can also define a reper. tory name, for example, "Call Harvey."

The heart of the dialer comprises fixed point DSPs, a ROM-based design particu larly suited for cellular phones. The DS has a number of built-in hardware fea tures that speed the implementation c speech recognition algorithms. Con sequently, the phones make full use c state-of-the-art digital technology t. maximize available telephone channe bandwidth.

Voice dialing features can be added t cellular telephones by simply increasin system memory — other DSP devices ar not required. The single speech codin DSP can be time shared to handle voic recognition as well because both func tions do not need to run simultaneously Further, integrated cellular telephone can use the same DSP to control othe functions, such as vehicle entertainmen equipment, climate, and windshiel wipers.

Voice dialer ROM and RAM com binations can be varied to handle differ ent size boot programs, program memor and data. The programs differ dependin on the number of telephony applicatio and the functions provided. An analo interface to the telephone handset, a alpha-numeric display, and interrupt driven connections to the telephor handset complete the set-up.

New product development

To aid in the design of new speech re ognition products, the dialer doubles as development system. An RS-232 inte face, for example, supports download external software and provides a condu for control and input information other systems associated with the diale As a result, the voice dialer is easily int grated into a specific application en ronment or another development syste and evaluated.

The RS-232 port downloads to a sep rate 64k RAM in the voice dialer. The bo program transfers the downloaded pr gram and data to the correct DSP memo

The dialer has uses other than the phone application. They include perso computers or workstations where vo

EVERYTHING OLD IS NEW AGAIN

Speech recognition technology is not new. A speaker verification system for military security was introduced in 1974, several years after research began in the 1960s. Even then, the system was said to be superior to fingerprint identification. TI also used a version of the system to control entry to its own computer center.

Today, speech and development systems are designed for a variety of applications, including text-to-speech, record/playback, telephone management, language recognition and speaker verification. Also, credit card verification systems are now widely used.

Text-to-speech algorithms convert ASCII text (as it appears on a computer monitor) into spoken English. The computer-generated voice is natural, intelligible, and has an unlimited vocabulary. Specific applications include inventory assessment, order entry input, and status review.

Record/playback applications are similar to tape recorders or dictation machines. The user can record notes, speeches and other material. However, computer storage provides greater clarity than magnetic recordings and enables the recorded file to be easily merged with other data files.

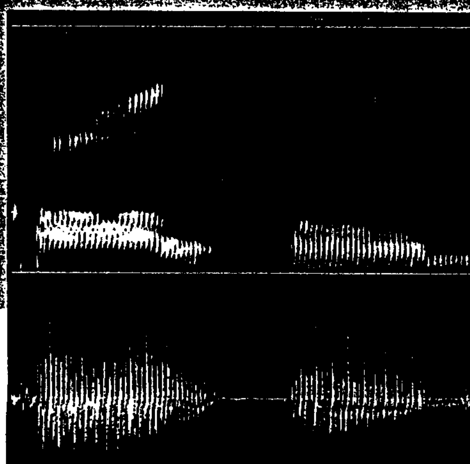
Telephone management systems employ computers to answer telephones, replay messages, and dial other telephones. Applications can be more complex than simple voice mail. In computer banking, for example, customer transactions phoned in can be confirmed at each step of the process by a synthesized voice.

Language recognition enables a computer to recognize complete sentences as they are spoken. One system, for example, can handle applications requiring up to 2,000 words. Language recognition goes beyond mere word recognition; entire sentences are analyzed using context analysis to help determine what is spoken. Language recognition is particularly useful in applications where keyboards cannot be used.

Speaker verification identifies a person through his or her unique voice characteristics. As such, it is ideal for a wide variety of security or entry control applications.

Successful applications arise from a melding of research and development in speech and semiconductor technology, and speech algorithms. For example, established multiple speech databases help create speaker-independent models for the digits used in the voice dialer.

Speech application development requires special software and hardware tools and utilities, and run-time libraries. Such software is available for a variety of DOS and Unix platforms. For example, a speech system tool kit (Speech System V) is available for Xenix or Unix systems running on Intel 80386-based computers. The tool kit also contains an interface for Unix systems operating on minicomputers.



DSP algorithms recognize the digitized form of an analog speech pattern. The top waveshape is a speechgraph of the words call home. The lower waveform is a spectrogram of the same phrase.

SS 04421

mory, so either eak to a car tele office" or "Cal lso state the num g the words zer . or the word "oh lso define a reper , "Call Harvey." r comprises fixed d design particu phones. The DS -in hardware fea mplementation o lgorithms. Con make full use o al technology t lephone channe

s can be added t simply increasing er DSP devices ar gle speech coding d to handle voic cause both func n simultaneously ellular telephone to trol other cle rtainment and windshield

and RAM com l to handle differ program memory s differ dependin hony application vided. An analo hone handset, a y, and interrupt o the telephone et-up.

evelopment

of new speech rec dialer doubles as An RS-232 inter ports downloade provides a condui t information t ed with the dialer ialer is easily int application env velopment syste

wnloads to a sep ce dialer. The boo d oaded pr rrect DSP memor other than the c y include person tions where voi

recognition is used instead of keyboard input. Also, voice input can supplement factory automation and process inspection data for various machines and computers.

A speech recognition system can also provide hands-off control of a vehicle entertainment system, climate control, windows, windshield wipers, and door locks. For example, a driver can select a radio station with his voice or change the interior temperature without removing his hands from the steering wheel. The voice system can also query the vehicle for fuel status and mpg ratings. Even more elegant features can be had at negligible cost, such as a voice lock that allows the vehicle to be started only by authorized persons.

A demonstration voice dialer system is contained in a portable, briefcase-size box. It is powered by either a 220/110-Vac supply or 12 Vdc through a vehicle cigar lighter receptacle. Such a portable voice dialer can be used as a development system or a test set to diagnose faults in telephones in other mobile units.

The voice dialer circuit is located on one printed-circuit board with programmable array logic (PAL) to minimize the number of individual support logic chips. Voice dialer subsystems include analog circuits and codec, processor and RAM memory, processor control and EEPROMS, display and communications port, and power.

Application-specific grammar

An algorithm can be loaded that makes the dialer recognize up to 25 words without discriminating male or female voices. And application-specific grammar can be either downloaded to the system through an RS-232 serial port, or installed at the factory.

A grammar is also called a sentence model. The DSP and speech recognition algorithms understand and respond to sentence models, and control the syntax by which the words are put together.

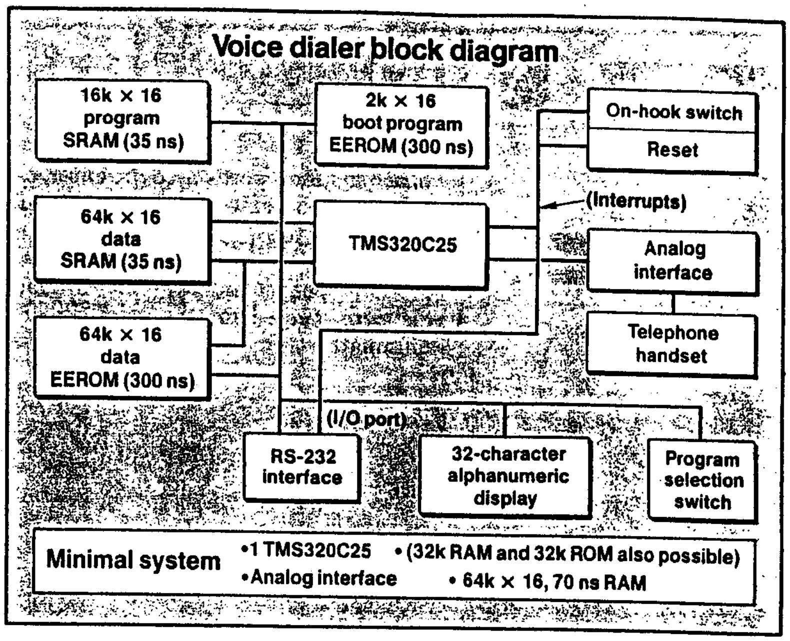
After the grammar is loaded, the voice dialer recognizes the following sequence of commands spoken in any order; call office, call home, or number (digits).

In this sequence, number is a digit string of any length, for example, number 516-666-7777 is a legal sentence. A 1-s pause (or other adjustable value) terminates any speech. When the voice dialer recognizes a complete phrase followed by the pause, it displays a period (.) on the voice dialer 32-character alpha-numeric liquid-crystal display screen. The commands 'enter' or 'cancel' can also terminate the connection.

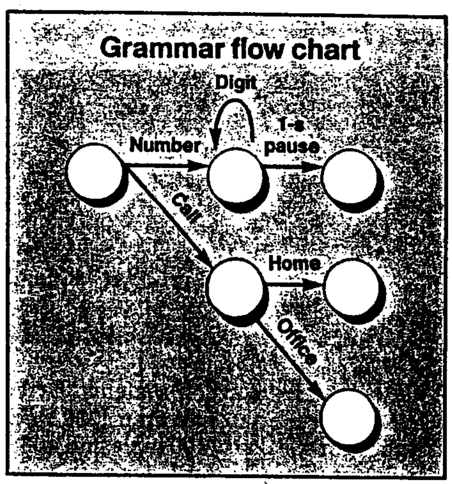
Pressing the off-hook switch on either the voice dialer case or the handset restarts the voice recognition process. In fact, the system recognizes just one command each time the phone goes off hook.

Other application grammars also are possible. An application may, for example, require that the speech recognition system recognize names and the word call as in the command call John Jones.

A basic voice dialer vocabulary consists of 11 digits (zero through nine and the word oh for zero) and four words (call, office, home, and number). But other words are easily added to the application grammar. In one version of the dialer,



The voice dialer requires either a TMS320C25 or TMS320C51 DSP with data memory, program memory, and EEPROM. A telephone handset interface, RS-232 port, display, and various switches comprise a system with a digital configuration that is different for each speech recognition algorithm that it employs.



Flow chart shows operation of the voice dialer when application-specific grammar is loaded. Here, the commands call office, call home, and number (digits) are possible, where digits is a digit string of any length.

SS 04422

TARGET TOPIC

AUTOMOTIVE

other common words used are enter, cancel, area, code, extension, and emergency.

The database connection

Many speaker-independent word models were created for the voice dialer to eliminate a training phase needed by earlier speech recognition systems. By collecting speech samples from 200 native American speakers (100 male and 100 female), statistical models for each vocabulary word were created. Thus, the likelihood of an unrecognizable word was largely diminished. Care was taken to sample different geographical regions to reflect various dialects. The repertory of voice information is archived in a speech database.

Recognizing that different accents need to be accommodated in certain applications, a speaker-adaptive operating mode was developed. In this mode, the voice dialer boots up with a speak er-independent model. The model is "seed" and the voice dialer controlling al gorithm continuously adapts the mod to the user in what is called a voice diale training mode.

Many novel applications also are poss ble using the database concept. For exam ple, a vocabulary may be developed that specific to one automobile manufactu or customer. For some applications, su as a personalized car phone that is di abled when others try to use it, TI c supply speaker-dependent capability f a code word.

In the present voice dialer, all need voice recognition functions, such as HM algorithms, signal processing, and gra mar control are performed by one D For more complex applications, howe such as large vocabularies and more co plex grammars, more than one DSP m be needed. Multiprocessor architect allows algorithm partitioning so t larger vocabularies may be recogni and accommodated.

Experimental versions of a mul processor DSP architecture for speech r ognition have already been made. many as 32 DSPs were connected which present, uses an IBM AT computer a host for development and input/outp functions. ■

DSP TARGETED FOR SPEECH RECOGNITION

The newest DSP, the TMS320C51, has an architecture especially configured for speech processing. The design speeds speech algorithm processing much as a hardware multiplier/accumulator speeds more conventional DSP signal processing.

For example, an important calculation performed by a speech algorithm is selecting a maximum or minimum value out of a set of values. Recognizing this, designers implemented such a maximum and minimum instruction set in hardware for the TMS320C51. A description of the maximum value instruction that compares only two values is illustrated here to help understand the more complex operation for a set of values.

Assume that the maximum value of two numbers is to be found. One is placed in the TMS320C51 accumulator, the other is placed in the accumulator buffer. The instruction CBGT initiates the following sequence: The contents of the accumulator are compared to the contents of the accumulator buffer, and the larger (signed) value is loaded into both registers.

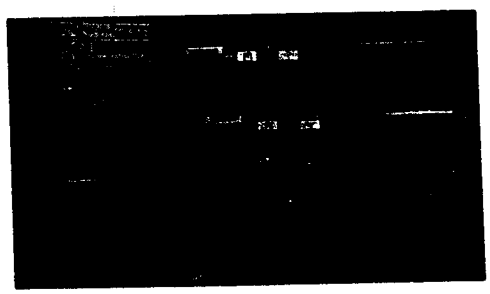
A carry bit is modified according to the comparison result. For example, if the contents of the accumulator are greater than or equal to the contents of the accumulator buffer, the carry bit is set to 1; otherwise it is zero. A similar procedure finds the least value in a set of values using the CRLT instruction.

A hardware feature of the TMS320C51 that makes it particu larly suited to voice recognition is that, unlike other DSPs, the C51 performs single-cycle 16 × 16-bit multiplications in 35 to 50 ns. Data shifting and address manipulations also are in hardware rather than microcode or software.

Speech recognition algorithms typically are arithmetic intensive and need to access as much DSP power as possible. The C51 DSP features a zero overhead context switch on inter rupts. That means no extra cycle time is needed to save d restore data when an interrupt is received. Because no timing cycles are used for data save/restore, that time is available fo computation.

The TMS320C51 is a fifth-generation digital signal process and a fixed-point machine. Available in a 132-pin Quad-Fla Pack package, the 5-V static CMOS Harvard architectu (separate data and program buses) DSP can be tested usin the industry standard JTAG IEEE P1149.1 boundary scan logi

Capable of more than 28 Mips, the DSP features on-chi ROM, program/data RAM, dual-access data RAM, and memor security. Also on-chip are address-mapped software wai state generators, serial ports, a hardware timer, five intern and four external user-maskable interrupts, and 64k I/O por accessed by sixteen 16-bit address lines.



Texas Instruments Speech System V Toolkit is a software development package used with a 80386-based computer to create speech programs. The tool kit provides the environment to make systems for voice recognition, record-and-play, text-to-speech, and telephone management. An option is also available for speaker verification applications in security products.